

UNIVERSITÉ DE MONTRÉAL



CENTRE DE RECHERCHE SUR LES TRANSPORTS



CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

SIX NOTIONS OF EQUILIBRIUM AND
THEIR IMPLICATIONS FOR TRAVEL MODELLING
EXAMINED IN AN AGGREGATE DIRECT DEMAND FRAMEWORK

by
Marc Gaudry

Forthcoming in Behavioural Travel Modelling
edited by Hensher and Stopher
Croom Helm, London 1978.

Final draft of discussion paper presented at
the Equilibrium workshop of the Third International
Conference on Behavioural Travel Modelling 1977
held at Tanunda, South Australia.
Comments welcomed.

This research was made possible by a grant from the
Research and Development Center of Transport Canada

Université de Montréal

Cahier # 7613 - Département de sciences économiques
Publication # 51 - Centre de recherche sur les transports
Juin 1977

Dépôt légal - 1er trimestre 1978

ABSTRACT

This paper focuses on notions of equilibrium used in the formulation and estimation of models which explain observed amounts of travel in terms of supply and demand considerations. The nature of services involved in static and dynamic equilibrium is examined; problems of specification of derived demand and income equilibrium are explored; the formulation of equilibrium in situations of producer-consumer interaction and the interface between the money or time trip inputs and network infrastructures are analysed.

RÉSUMÉ

Cet article présente six notions d'équilibre utilisées pour formuler les modèles qui expliquent les déplacements des voyageurs par des structures d'offre ou de demande. Après un examen de la nature des services de transport, les problèmes de formulation d'équilibres pour des demandes intermédiaires sont abordés, ainsi que ceux qui sont reliés aux effets de revenu; la distinction entre équilibre du marché et équilibre du réseau est présentée. Des notions simples d'équilibre dynamique sont effleurées.

SUMMARY

INTRODUCTION	1
THE NATURE OF TRAVEL OUTPUT.	2
The Demand Side: Dimensions of Transit Trips	2
The standard format.	3
The joint consumption format	3
Extension to Other Modes and to the Supply Side	5
EQUILIBRIUM FOR DERIVED DEMANDS.	5
With Variable Activity Levels	6
Work related trips	6
Trips for other purposes	8
With Fixed Activity Levels.	9
Implications.	9
EQUILIBRIUM WITH RESPECT TO INCOME	10
General Considerations.	10
Specific Considerations	11
Implications.	12
MARKET EQUILIBRIUM	12
The Car Trip Market	13
The Bus Trip Market	15
Implications.	16
Market clearing: the existence of equilibrium	16
Identification: the structure of equilibrium.	16
Simultaneity and pure structural relationships	17
Other stochastic specification problems.	17
NETWORK EQUILIBRIUM.	18
The Assumption of Generalized Itinerary Cost Minimization	18
The Network Equilibrium Problem	21
Implications.	21
Linearity in the form of the generalized cost function	21
Market equilibrium and network equilibrium	22
Demand estimation and the true itinerary input vector.	22
DYNAMIC EQUILIBRIUM.	24
CONCLUSION	25
APPENDIX A	25
APPENDIX B	26
BIBLIOGRAPHY	26

SIX NOTIONS OF EQUILIBRIUM AND THEIR IMPLICATIONS FOR TRAVEL MODELLING EXAMINED IN AN AGGREGATE DIRECT DEMAND FRAMEWORK

Marc Gaudry*
Université de Montréal, Montréal, Canada

INTRODUCTION

This paper purports to analyse the implications of various notions of equilibrium used in the formulation and estimation of transportation models which explain observed amounts of travel in terms of supply and demand considerations.

We will perform this analysis by spatialising in increasing details an aggregate direct demand function for trips in an area during period t :

$$T_t^d = f^d(S_T, PD_T, P_v, P_f, A_f)_t + e_t^d, \quad (1)$$

where the quantity of trips demanded depends on the service level of transport modes S_T , on the price of trip distance units PD_T , on the prices of commodities (including labour) bought or sold in variable and fixed amounts, P_v and P_f , and on the amount of fixed commodity trades A_f . For the sake of convenience, time subscripts on all variables and on the stochastic term will be dropped in the first five sections of the paper in which static notions of equilibrium are presented and equilibria are assumed to occur within this implicit time period; the time subscripts will be reintroduced in the sixth section where time paths of adjustment of variables are explicitly considered. We assume throughout that the location (but not the level) of activities is invariant. These time period, level of aggregation and land use assumptions are meant to isolate a number of situations in which short-run equilibrium notions have relevance to transportation modelling. We shall not address directly equilibrium situations in which the time span required for adjustments to occur is very large, such as those of interaction between transportation

*This paper was made possible by a grant from the Research and Development Centre of Transport Canada and benefited from comments on earlier drafts by Earl Ruiter and by Michael Florian with whom fruitful discussions on the nature of network equilibrium were also frequently held. Michael Wills gracefully permitted the reference to his forthcoming thesis.

networks and land use. Our analysis is meant as a methodological document focussed on six equilibrium issues and not as a review of the literature of equilibrium concepts and practices.

The first section will contain an examination of the nature and dimensions of the "commodity" involved in presumed equilibria and explore the incorporation of all of these dimensions in the formulation of the dependent variable of travel demand functions. The second section will show that the use of activity level variables to explain derived trip levels introduces specification problems which might go as far as lead one to expect negative correlations between observed trip levels and activity levels in certain circumstances. The third section argues for greater understanding of money and time income effects in trip analysis. The fourth and fifth sections explore the distinct and interrelated notions of market and network equilibrium in the joint determination of travel flows and service levels. The final section introduces dynamic considerations. We shall therefore examine successively the role of each type of variable used in Equation (1). The notions of disequilibrium and the estimation problems discussed at various occasions will not be regrouped in specific sections.

THE NATURE OF TRAVEL OUTPUT

The first issue which naturally comes to mind is that of the nature of the commodity which is the object of equilibrium, or of the dimension of the dependent variable in Equation (1). We wish to argue that better use might be made of these dimensions in the formulation of models. To that end, let us specify the dimensions of transit trips in our prototype demand equation.

The Demand Side: Dimensions of Transit Trips

Consider an individual living in an urban area and facing a network of bus lines of identical frequency S_{1b} and speed S_{2b} .

Assume that the i^{th} trip using one or more lines consists of waiting time t_{1bi} , in-vehicle time t_{2bi} and seat-mileage t_{3bi} .

Under uniform price per mile and service level conditions, the consumer's total demand for trip characteristics corresponding to T_b trips can be expressed as

$$(T_{1b} \ T_{2b} \ T_{3b})^d \equiv T_b^d = f_b^d(S_{1b}, S_{2b}, PD_b, \text{ETC}) + e_b^d \quad (2)$$

where total seat-waiting time $T_{1b} = \sum_i t_{1bi}$, total seat-travel time $T_{2b} = \sum_i t_{2bi}$ and total seat-mileage $T_{3b} = \sum_i t_{3bi}$.

In this perspective, changes in the level of independent variables may induce the consumer to change the number of trip T_b^d , their length d_i or both: a higher bus speed S_{2b} could leave trip demand T_b^d constant but affect the consumption of trip characteristics in one of various ways (e.g. longer trips (T_{3b} increases)

and more or less in-vehicle time T_{2b}). In Equation (2), the frequency and speed properties of the mode behave as time prices and say nothing about actual trip characteristic purchases per trip. In real contexts, price and service levels are not identical on all lines. Moreover, Equation (2) does not tell us the location of the demands. A number of approaches can be used to circumvent these problems.

The standard format. The first one consists of disaggregating trips by origin-destination pair. This makes two simplifications possible. Firstly, the vector $(S_{1b} S_{2b} PD_b)$ can be transformed into a vector of trip rates and Equation (2) can be written as

$$\begin{aligned} (T_{1b} T_{2b} T_{3b})_{ij}^d &= [(K_{1b} K_{2b} K_{3b})_{ij} \cdot T_{bij}^d] \\ &= f_b^d [(K_{1b}, K_{2b}, P_b)_{ij}, ETC_{ij}] + e_{bij}^d, \end{aligned} \quad (3)$$

where P_b denotes the price for K_{3b} units of distance on buses. It is important to realize that, even if the transformation results in waiting time K_{1b} and travel time K_{2b} per trip being used on both sides of the equation, the significance of these variables differs: on the left-hand side, they enter as commodities; on the right-hand side, they enter as resources or prices.

Secondly, the commodity-vector of demanded characteristics can be neglected and only T_{bij}^d kept as dependent variable, as in

$$T_{bij}^d = f_b^d(K_{1b}, K_{2b}, P_b, ETC)_{ij} + e_{bij}^d. \quad (4)$$

This second simplification can blur even further the distinction between commodity and resource. Utility is defined over $(K_{1b} K_{2b} K_{3b}) \cdot T_{bij}$, a vector of commodities, and not over (K_{1b}, K_{2b}, P_b) ,

which consists of prices. Appendix A of a Charles River Associates report (Ref. 5) makes this point clearly. By contrast, it is not clear in the current rationale of the logit model (Ref. 8), which does not use an indirect utility function, what is a price and belongs to a budget constraint and what is a commodity and belongs to the utility function.

The joint consumption format. Let us develop an alternative approach based on the simple idea that all of the constitutive elements of trips should be used as dependent variable. Consider that the trip elements cannot be bought separately but have to be bought jointly; rewrite Equation (3), dropping $i j$ subscripts, as

$$\begin{aligned}
 J_b^d &= (\gamma_1 T_{1b} + \gamma_2 T_{2b} + \gamma_3 T_{3b})^d \\
 &= f_b^d (K_{1b}, K_{2b}, P_b, \text{ETC}) + e_b^d,
 \end{aligned}
 \tag{5}$$

where the trip elements are weighted and the γ weights sum to one. In this perspective, the individual's demand curve for J_b results from the vertical addition of demand curves for each of the trip component elements (Ref. 16) in the spirit of the welfare conditions for jointly supplied goods (Ref. 27). The aggregate demand curve is obtained by conventional horizontal addition over agents.

We recently compared transit demand elasticities obtained from a standard specification of the dependent variable, as in (4), and from a joint consumption specification, as in (5). We did not have disaggregate individual trip data and consequently made considerable errors of observation on the components of the dependent variable; the results for selected transit trip demand elasticities listed in Table 1 supported the standard specification because the in-transit time demand elasticity had an unreasonably high and negative value under the joint consumption specification.

TABLE 1 Selected Transit Demand Elasticities Derived from Different Specifications of the Dependent Variable for Montreal Adults*

	Standard	Joint Consumption
Bus-metro fare	.18 (- 6.41)	.25 (- 0.60)
waiting time	.16 (- 2.15)	.79 (- 13.02)
in-transit time	.49 (- 12.69)	-1.18 (- 16.99)
Car in-transit time	-.07 (0.41)	.10 (- 0.18)
Income	-.52 (3.67)	-0.77 (0.63)

It was interesting to note that the weights (and t-ratios) of the components of the dependent variable were: -.14 (-13.90) for waiting time, -.02 (-18.43) for in-transit time and 1.16 for mileage; the calculated value for J_b^d was negative which had the prima facie implication that derived demands yield disutility and would lead one to expect that, for a final good, J_b would be positive. These results are interesting enough to suggest that the joint consumption specification should be tested with a better data set. If trips really consist of a vector of commodities, there should be a way of entering these elements as dependent variables and of using that information directly; at present, the standard format of the dependent variable can be seen as a special case of the joint consumption formulation in which

*Numbers in parentheses are t-ratios and give the sign of the appropriate independent variable. More details on model formulation and estimation are given in Appendix A.

$(\gamma_1 K_1 + \gamma_2 K_2 + \gamma_3 K_3)$ is constrained to equal one.

Extensions to Other Modes and to the Supply Side

The vector of characteristics demanded from i to j can also be written for automobile trips:

$$(T_{1a} \ T_{2a} \ T_{3a})_{ij}^d, \quad (6)$$

where elements respectively denote seat-waiting time, seat-travel time and seat-mileage.

This definition of the dimension of trips is also applicable to the supply side; the quantity supplied is

$$(T_{1m} \ T_{2m} \ T_{3m})_{ij}^s = [(K_{1m} \ K_{2m} \ K_{3m}) \cdot T_m]_{ij}^s, \quad (7)$$

where m denotes any mode and the elements denote the number of seats provided, T_m , multiplied by their waiting time K_{1m} , by their travel time K_{2m} and by the distance K_{3m} they cover in going from i to j . Other dimensions (comfort, safety, walk time, etc.) could in principle be added in similar fashion both on the demand and on the supply side. Whether they are used as dependent variables in supply equations or as explanatory variables in demand equations, the elements K_{1m} and K_{2m} are of course transformations of frequency and speed variables into time per trip variables.

For the purposes of the following section, we shall use the standard format, T_m , of the dependent variable in demand functions.

EQUILIBRIUM FOR DERIVED DEMANDS

A second issue of importance in the explanation of the amounts of trip making is that of the presence of activity level variables in trip generation and direct demand equations. These variables introduce the overall level of economic activity as an explanatory variable of the frequency of trips. A paradigm equation could be

$$T^d = f^d (K_{1T}, K_{2T}, P_T, P_A, A) \quad (8)$$

in which the money and time prices of travel, the price of other consumption activities and the level of these activities appear as arguments. Functions of that general form are normally estimated on the assumption that activity levels $A = (A_1, \dots, A_n)$ are exogeneous and predetermined. Yet the underlying rationale for this procedure is far from clear. We shall argue that the presence of activity variables in numerous trip demand equations often introduces a simultaneous equations bias and can sometimes be inappropriate. Let us examine situations which may increase our understanding of the problem.

With Variable Activity Levels

Certain classes of cases arise if the levels of various trades for labour and other goods are allowed to be determined by the prices of these activities and by travel prices. Work and shopping trips are examples.

Work related trips. Consider the following simplistic version of the structure of an aggregate market for labour. The demand function for units of labour services L (say man-hours) is simply

$$L^d = \alpha_1 P_L + \alpha_2 \text{OTHER}_d, \quad (9)$$

where P_L denotes the price of labour, OTHER_d is self explanatory and the α 's are parameters. Specify the supply function as

$$L^s + \beta_1 T^s = \beta_2 P_L + \beta_3 P_T + \beta_4 \text{OTHER}_s, \quad (10)$$

where T^s denotes the amount of travel workers are willing to do as they supply L^s work services and P_T is the price of travel.

In this formulation L^s and T^s are jointly supplied and for that reason are jointly dependent variables. This formulation incorporates the fact that working involves supplying both man-hours on the job and work-related trips, elements which have certain relative weights (here the parameter of the first element has been "normalized" or constrained to be equal to one) in the decision to supply one unit of "working". These jointly supplied elements are made a function both of P_L and of P_T because there is no reason to suppose that one earned dollar has the same effect as one dollar spent on travel (if one believed otherwise, one would simply write $\beta_2 = \beta_3$) on the supply of these jointly produced elements.

The market clearing equation is

$$L^s = L^d. \quad (11)$$

If the amount of travel per unit of labour is not technologically fixed and if the labour market is in equilibrium, one can write the reduced form equation* for the jointly dependent pair of variables as

$$L + \beta_1 T^s = \frac{\beta_2}{\alpha_1} L + \beta_3 P_T + \beta_4 \text{OTHER}_s - \frac{\alpha_2 \beta_2}{\alpha_1} \text{OTHER}_d, \quad (12)$$

*In practice one would redefine $L^s + \beta_1 T^s$ as W and add that identity to the system consisting of the demand, supply and market clearing equations.

or as

$$(1 - \frac{\beta_2}{\alpha_1})L + \beta_1 T^S = \beta_3 P_T + \beta_4 \text{OTHER}_S - \frac{\alpha_2 \beta_2}{\alpha_1} \text{OTHER}_D . \quad (13)$$

To obtain a derived demand function for work trips, one simply relabels T^S as T^D and collects terms:

$$T^D = \frac{\beta_2 - \alpha_1}{\alpha_1 \beta_1} L + \frac{\beta_3}{\beta_1} P_T + \frac{\beta_4}{\beta_1} \text{OTHER}_S - \frac{\alpha_2 \beta_2}{\alpha_1 \beta_1} \text{OTHER}_D . \quad (14)$$

We note the following

- i) because L , an endogeneous variable of the system, appears on the right hand side of Equation (14), an ordinary least-squares regression of T on L , P_T and OTHER would produce biased estimates of the resulting parameters;
- ii) if these parameters are called γ 's and Equation (14) is rewritten as

$$T^D = \gamma_1 L + \gamma_2 P_T + \gamma_3 \text{OTHER} , \quad (15)$$

the sign of γ_1 should be negative (given $\alpha_1 < 0$; $\beta_1, \beta_2 > 0$): the fact that one expects the parameter of the employment level to be positive in such equations indicates the presence of an unresolved problem. By contrast, the sign of a variable included in both OTHER_S and OTHER_D will (given $\alpha_2 < 0$; $\beta_2 > 0$) depend on the size of each component of

$$\gamma_3 = \frac{\beta_4}{\beta_1} - \frac{\alpha_2 \beta_2}{\alpha_1 \beta_1} ;$$

- iii) although not all work trip generation equations have the appearance of Equation (15), many include its basic elements. This is at it should be: the variable P_T , for one, is important in the decision to supply work. Indeed the changes in the labour participation rate which accompany the business cycle are certainly partly due to the fact that a decrease (or an increase) in the number of jobs typically implies longer (or shorter) work trips to similar jobs. In fact such changes in the participation rate, notably of women, may give an important clue concerning the numerical importance of increases in the supply of labour which result from transportation network improvements.

By contrast, if the amount of work trips per unit of labour is technologically fixed as in

$$T^S = kL^S , \quad (16)$$

where k is a technological constant, then the trip demand equation is

$$T^d = kL. \quad (17)$$

Under this rigid view, it would be inappropriate to use as a trip generation equation expression (14) or expression (15) because the former can be rewritten from (17) as

$$L + \beta_1 kL = \frac{\beta_2}{\alpha_1} L + \beta_3 P_T + \beta_4 \text{OTHER}_S - \frac{\alpha_2 \beta_2}{\alpha_1} \text{OTHER}_d, \quad (18)$$

which implies that a regression of T on L , P_T and OTHER would in effect be, up to a normalization factor k , a regression of L on itself:

$$T^d = kL = \frac{\beta_2 - \alpha_1}{\alpha_1 \beta_1} L + \frac{\beta_3}{\beta_1} P_T + \frac{\beta_4}{\beta_1} \text{OTHER}_S - \frac{\alpha_2 \beta_2}{\alpha_1 \beta_1} \text{OTHER}_d. \quad (19)$$

The proper trip generation equation is either (17) or the relationship obtained after substituting in (17) for the value of L derived from (19):

$$T^d = kL = k \left[\frac{\beta_3}{\beta_1} \delta P_T + \frac{\beta_4}{\beta_1} \delta \text{OTHER}_S - \frac{\alpha_2 \beta_2}{\alpha_1 \beta_1} \delta \text{OTHER}_d \right], \quad (20)$$

where $\delta = \frac{\alpha_1 \beta_1}{\alpha_1 \beta_1 k - \beta_2 + \alpha_1}.$

It is clear that a regression of T on P_T and OTHER would be a meaningful trip generation equation; it does not include the activity level variable L explicitly.

Trips for other purposes. Similar reasoning can be applied to trips for purposes other than work. If R^S and R^d denote the amount of goods supplied and demanded in retail establishments, the supply, demand and market clearing equations are

$$R^S = \alpha_1 P_R + \alpha_2 \text{OTHER}_S, \quad (21)$$

$$R^d + \beta_1 T^d = \beta_2 P_R + \beta_3 P_T + \beta_4 \text{OTHER}_d, \quad (22)$$

$$R^S = R^d. \quad (23)$$

Again, Equation (22) allows the price of retail goods and the cost of travel to have distinct effects on R^d and T^d which are bought jointly in varying proportions. As for (14), regressing shopping trips on retail sales, the cost of travel and other variables would yield biased estimates of mixtures of structural parameters. Analogs of (17) or (20) would be appropriate.

The presence of changing activity level variables in trip demand

equations is therefore by no means unambiguous because these equations are in fact summary descriptors of more complex underlying equilibria. At the best, their presence introduces a simultaneous equations bias in equations of form (15): the strength of the bias depends on the constancy of coefficient k ; at the worst, that form is tantamount to a regression of the activity variable on itself and on other variables: the proper format is then either (17) or, going a step behind, (20). The latter form includes as regressors the prices of travel and non travel goods (under OTHER_s).

With Fixed Activity Levels

Suppose that the level of activity A is fixed. In the presence of a technological relationship between the level of A and trip making, the derived demand for A -linked trips is given by an analog of Equation (17):

$$T^d = k A + e^d \quad (24)$$

and there is no analog of Equation (20) which would explain the level of A . The aggregate demand for trips is given by the function

$$T^d = k A + f^d (P_T, P_A, \text{Other prices}) + e^d, \quad (25)$$

where P_A is in effect a measure of income if A denotes work activity and a measure of disposable or discretionary income after fixed commitments are met if it denotes some other fixed activity. If there is no constant relationship between activity levels and trip making, Equation (25) reduces to

$$T^d = f^d (P_T, P_A, \text{Other prices}) + e^d \quad (26)$$

in which activity level variables are absent. Equations (25) and (26) are perfectly consistent with the general demand function derived in the Appendix of the above mentioned Charles River Associates report (Ref. 5).

Implications

Exceeding care should then be taken with the inclusion of activity level variables in trip generation and direct demand equations. Further work is clearly needed to circumscribe the situations in which these variables may be included with reasonable expectations as to their sign and expected role. It may be found that travel prices influence the number of trips per unit of activity for work activity, but not the level of activity itself and that, for other purposes, they influence both trip inputs per unit of activity and the level of activity. If it is found that some significant interdependence exists between activity variables and some of the travel prices, it may be appropriate to use estimation methods which reestablish this independence, at least asymptotically. In the meantime, if one uses equations like (24)

or (26) to explain trip making for various purposes or by various groups, it may be useful to estimate these sets of generation equations in grouped fashion in order to use, à la Zellner (Ref. 30), at least the information contained in the covariances of error terms across equations: the stochastic errors of work, shopping and social trip generation equations are certainly often correlated and the most efficient estimates should be obtained with Zellner's method.

The explicit treatment of most of travel demands as derived demands in a general equilibrium context should progress as macroeconomic demand models gradually allow for explicit transportation sectors (e.g. Ref. 21) and as systems of microeconomic equations, such as those of the Rotterdam model (Ref. 1) make efforts in the same direction. Until these approaches have developed further, adjustments must be made in ad hoc and partial fashion: the correct relationships among activity levels and travel must be specified by relying more on direct intuitive formulations in a partial equilibrium framework than on formulations strictly derived from rigorous general equilibrium approaches; the market balance is the underlying principle of the mechanism which the incorporation of activity level variables in trip frequency equations describes and one must be conscious of that in using simple formulations such as (8).

EQUILIBRIUM WITH RESPECT TO INCOME

A third issue in the explanation of travel behaviour is that of the role of income, the exact impact of which remains somewhat elusive due to the difficulty of the problem. We shall argue that fundamental work is needed in this area and that the current emphasis on mode split in the literature gives a distorted picture of time and money income effects.

General Considerations

Assume that consumer tastes are given. In standard demand analysis, it is traditional to distinguish between gross and net substitution (or complementary) and income effects as follows: if a demand function is written in general form as

$$T^d = f^d (P_T, P_{NT}, Y) , \quad (27)$$

where the prices of travel and non-travel goods and income are included, the gross substitution (or complementarity) effects are defined as $\partial T^d / \partial P_T$ and $\partial T^d / \partial P_{NT}$ and the gross income effect is defined as $\partial T^d / \partial Y$. The net effects arise from a distinction between the part of a price change which leaves consumers as well off as before, SUB_{P_T} or $SUB_{P_{NT}}$ and the part which amounts to a change in income, ΔY_{P_T} or $\Delta Y_{P_{NT}}$. When the price-specific income effect is thought to be small (when the consumers' expenditure on the

good constitutes a small part of their total expenditures), the gross substitution ($\partial T^d / \partial P = SUB_P + \Delta Y_P$) and income ($\partial T^d / \partial Y$) effects, estimated as coefficients of price and income of a market demand function, should not differ much from the net effects SUB_P and $(\Delta Y_P + \partial T^d / \partial Y)$.

Travel demand functions have to be more complex than Equation (27) because of the importance of time prices; an analog of that equation could be

$$T^d = f^d (K_T, P_T, P_{NT}, Y, TB), \quad (28)$$

where K_T denotes time prices of travel and TB is the consumers' time budget. (Ref. 6) has examined some of the fundamental problems involved in obtaining appropriate formulations of Equation (28) and in deriving time prices from parameter estimates. Because travel time and money expenditures constitute a very significant proportion of consumers' total allocations, gross and net money or time effects might really differ. Fundamental work is needed to measure these income effects.

Specific Considerations

In practice it will be tempting to make strong assumptions on the utility function in order to derive testable hypotheses which contain a reasonable number of parameters or which constrain the functional form of demand models. This is a task which requires great care: one could not easily accept functional forms based on the assumption of additive separability of the utility function which implies (e.g. Ref. 19) that all goods are substitutes (none are complements) and that all are superior. Indeed, these properties would be strong if one were interested in aggregate trip generation by mode among pairs of zones and if one wanted to know whether two modes are complements and whether public transit is an inferior good: they would mean that an improvement of one mode could never lead to an increase in demand for another mode and that inferiority ($\partial T_m^d / \partial y < 0$) of any mode m would be ruled out by the rationale of the model.

On a more applied level, strong emphasis on modal choice at the expense of trip generation can give a misleading impression of the (gross) impact of a price or income change on the demand by mode. This can be seen most clearly in two-stage models which incorporate a generation and a modal share part as follows:

$$T_m^d = f_g (\bar{P}_T, \bar{K}_T, A, Y) \cdot f_s (P_T, K_T, Y), \quad (29)$$

where $\bar{P}_T$ and $\bar{K}_T$ denote average money and time costs of travel and P_T and K_T are vectors of money and time costs of the modes considered individually. Gross price and income effects are the sum of two elements

$$\frac{\partial T^d}{\partial X} = f_s \cdot \frac{\partial f_g}{\partial X} + f_g \cdot \frac{\partial f_s}{\partial X} \quad (30)$$

which must both be considered in demand analysis. Even without using the distinction between gross and net effects, these elements might have opposite signs and the computation of the total impact of income (or price) changes on the demand for a particular mode T_m^d should take into account the sign and size of the effect on modal split (if any) and of the effect on the total number of trips. It is interesting to note that some urban direct demand models for transit work trips (Ref. 7) or for transit trips for all purposes (References 20; 14; 10) suggest, in contrast with mode split models, that public transit is not an inferior good, results which may precisely be due to the fact that the explanation of the trip frequency decision is not isolated from the mode choice decision in these models and which imply that public transit is not necessarily doomed in a growing economy.

Implications

A first implication of these considerations is the importance of distinguishing in models between the role of income variables as market disaggregation categories which represent tastes and their use to estimate the effect of income proper. In the latter capacity, their absence from some trip generation equations is a specification error which biases the estimates of the parameters of the retained variables.

A second implication is the need for precise measures or substitution and income effects both for money and for time prices. This will require fundamental work on the development and estimation of functional forms which permit these measurements without excessively strong assumptions on the utility function, a very difficult task indeed.

MARKET EQUILIBRIUM

Observed trip levels do not depend only on consumers' demands: they also depend on the behaviour of suppliers of trips. To bring into focus the importance of supply, we shall use the notion of market equilibrium. Although nobody thinks that it is a perfect notion, most might agree that it is a useful notion. We shall apply it to two modes, starting with the most difficult one, call for a greater recognition of the importance of distinguishing producers and consumers in explanations of observed trip levels and point to a number of implications of the notion for model estimation.

In order to specify the notion of market equilibrium, we shall assume that there exist single non overlapping itineraries by automobile a and by bus b among all i, j pairs of zones under consideration. We shall assume further that there is no

congestion on these itineraries. These assumptions define unique money and time prices for each mode among all zone pairs. Illustrative market structures are then as follows.

The Car Trip Market

In the car trip market, one specification of the demand, supply and market clearing conditions could be, if all V_a cars have the same capacity c_a :

$$[(K_{1a} \ K_{2a} \ K_{3a}) \ c_a \ V_a]^d = f_a^d(K_{1a}, K_{2a}, P_a, K_{1b}, K_{2b}, P_b, \text{OTHER}_{da}) , \quad (31)$$

$$[(K_{1a} \ K_{2a} \ K_{3a}) \ c_a \ V_a]^s = f_a^s(K_{2a}, P_a, \text{OTHER}_{sa}) , \quad (32)$$

$$[(K_{1a} \ K_{2a} \ K_{3a}) \ c_a \ V_a]^d = [(K_{1a} \ K_{2a} \ K_{3a}) \ c_a \ V_a]^s . \quad (33)$$

With cars, the waiting time K_{1a} is difficult to define but we shall leave it in these functions. If the simpler standard format is used and the amount of seats demanded and supplied, T_a^s and T_a^d , are assumed equal to $c_a V_a^s$ and $c_a V_a^d$, the market clearing condition (33) can also be written as

$$(c_a V_a)^s = (c_a V_a)^d , \text{ or } T_a^s = T_a^d . \quad (34)$$

These are a number of ways of understanding these equations. If one is thinking about an interzone taxi market, P_a is paid out to the taxi firm and OTHER_{sa} includes vehicle and wage costs and other variables such as the type of regulation of the industry. In that case, the supply curves proper are given by the partial derivatives of the supply vector with respect to that market price $\partial[(K_{1a} \ K_{2a} \ K_{3a}) \ c_a \ V_a]^s / \partial P_a$: one would expect the seat-waiting time element to be constant [$\partial(K_{1a} \ c_a \ V_a)^s / \partial P_a = 0$] if, by definition, K_{1a} is inversely related to V_a^s , and the other two elements to be positive. Slower travel speeds which increase production costs would have no impact on the seat-waiting time and would shift the other two supply curves to the left. Under the very strict definition of market clearing given above, most taxi systems suffer from excess supply because there is usually little pooling of taxi trips under prevailing money and time prices.

If one is thinking about a private car trip market, the structure

can be applied either within households or for groups of households. Given a cost sharing formula for out-of-pocket costs, all members of households demand car trips but only members with a driver's licence supply them and an increase in the latter's supply following an increase in their remuneration defines a number of standard supply curves (one for each element of the vector, as above). Because these drivers are not financially compensated for their driving time, an alternative definition of the supply function can be obtained by defining a generalized cost GIC_{sa} as

$$GIC_{sa} = \lambda_1 R_a + \lambda_2 K_{2a}, \quad \lambda_1 < 0, \quad \lambda_2 > 0, \quad (35)$$

where R_a is the net financial compensation of drivers per trip. If one defines a similar notion for the demand side as

$$GIC_{da} = \theta_1 P_a + \theta_2 K_{1a} + \theta_3 K_{2a}, \quad \theta_1, \theta_2, \theta_3 > 0, \quad (36)$$

seat demand and supply curve graphs can be constructed as in Fig. 1. At GIC level 1, slightly more than half of the seats are occupied. A rise in out-of-pocket costs which would raise the GIC uniformly to level 2 (by raising P_a and lowering R_a in such a way that $\Delta\theta_1 P_a = \Delta\lambda_1 R_a$) would increase the car occupancy rate from AB/AC to $A'B'/A'C'$.

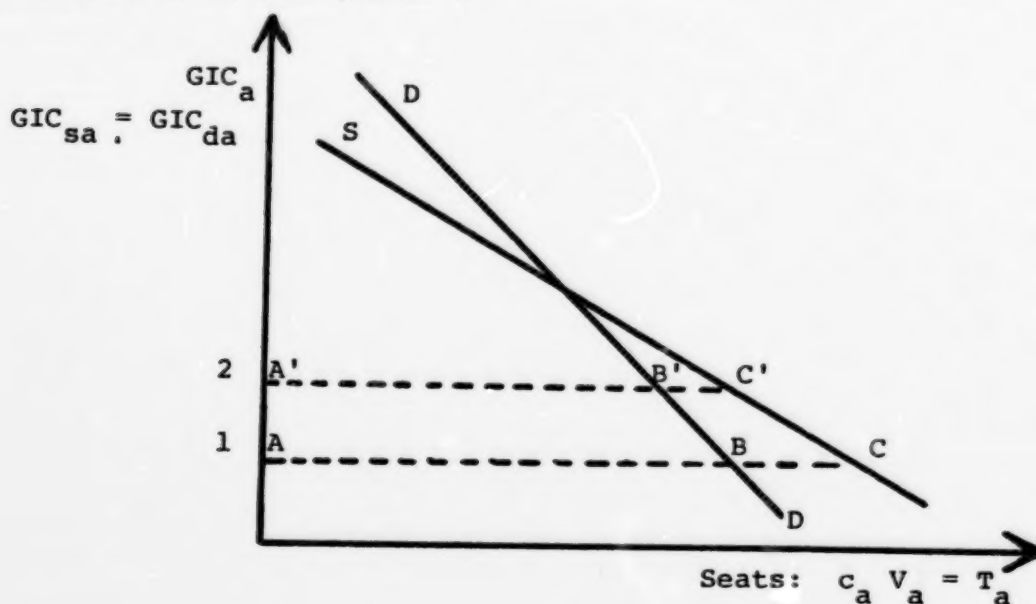


Fig. 1. Generalized itinerary costs and the car trip market

Although basic work is necessary to obtain convincing specifications of the car trip structural equations for various levels of aggregation of models, the distinction between the supply and demand sides of that market is the basis of the explanation of car

pooling, car occupancy rates and of the analysis of policy options which affect the supply side in a fashion distinct from the way in which they affect the demand side.

The Bus Trip Market

The structural equations for the bus market are:

$$[(K_{1b} \ K_{2b} \ K_{3b}) \ c_b \ v_b]^d = f_b^d (K_{1a}, K_{2a}, K_{1b}, K_{2b}, P_a, \tilde{P}_b, \text{OTHER}_{db}), \quad (37)$$

$$[(K_{1b} \ K_{2b} \ K_{3b}) \ c_b \ v_b]^s = f_b^s (\tilde{P}_b, C(K_{2b}, K_{3b}), \text{OBJ}_{bs}, \text{OTHER}_{bs}), \quad (38)$$

$$(c_b \ v_b)^s = (c_b \ v_b)^d, \text{ or } T_b^s = T_b^d. \quad (39)$$

These equations are the same as those presented in (Ref. 13) except for the dependent variable which is here presented in all of its dimensions instead of simply as capacity demanded or supplied.

We have assumed that a bus monopoly with certain objectives OBJ_{bs} and identical vehicles of capacity c_b supplies on each itinerary services the unit costs of which are a function of the speed and distance covered, $C(K_{2b} \ K_{3b})$, on that itinerary. The supply curves are obtained as usual by taking the partial derivative of the supply vector with respect to the regulated price $\tilde{P}_b$. With identical buses the waiting time will be inversely related to the number of vehicles and it is reasonable to presume that, if the firm's objective is to make zero profit, we should have

$$\partial (K_{1b} \ c_b \ v_b)^s / \partial \tilde{P}_b = 0, \ \partial (K_{2b} \ c_b \ v_b)^s / \partial \tilde{P}_b < 0 \text{ and}$$

$\partial (K_{3b} \ c_b \ v_b)^s / \partial \tilde{P}_b > 0$. An increase in wage costs would have no effect on the first element and shift the other two supply curves to the left. Strict market clearing occurs when the capacity supplied is strictly equal to the capacity demanded over a given itinerary.

In both the car and the bus trip markets, the classical supply curve has simply become a supply vector. The apparent paradox $\partial (K_{1a} \ c_a \ v_a)^s / \partial P_a, \ \partial (K_{1b} \ c_b \ v_b)^s / \partial \tilde{P}_b = 0$ results from the fact that we have expressed the frequency dimension of service in terms of a transformation of waiting-time; had we kept measures of frequency of service S_{1a} and S_{1b} , the paradox would not have

appeared.

The proper estimation of supply and demand functions for these two modes will require meeting a number of conditions which we shall presently examine.

Implications

Market clearing: the existence of equilibrium. In practice, estimation of the full model for each mode does not require that the strict market clearing conditions defined in Equations (34) and (39) hold. What must exist is market clearing in the weaker sense of

$$T_a^s > T_a^d, \text{ and } T_b^s > T_b^d, \quad (40)$$

otherwise the demand relationships might not be observed: if excess demand prevails, the amount supplied is observed but the amount demanded is not.

Because of the difficulty of formulating convincing structural equations for the car market, even this weak notion of market equilibrium is very approximate: in the car trip market, there is of course a difference between excess supply of seats and excess supply of vehicle trips; there is also a difference between aggregate and individual observations because excess supply in the aggregate is compatible with excess demand for individual cars. Yet the notion is useful because it implies that an aggregate number of observed trips by car does not represent the quantity truly demanded or supplied. This error of observation on the dependent vehicle becomes a part of the stochastic term associated with each car market relationship.

In the bus market, excess demand means that passengers have to wait for more than one vehicle to find space. If waiting time K_{1b} is measured from data on vehicle operations, errors of observation on both that variable and the dependent variable can occur. The use of sophisticated disequilibrium estimation techniques (Ref. 9; Refs. 25 and 26) might be required. Such situations might arise frequently if it does not pay the bus firm to supply adequate capacity on certain itineraries.

Identification: the structure of equilibrium. But the existence of an equilibrium is not sufficient for modelling. If, in each market, the list of variables included in the supply and demand functions were identical (and forgetting the presence of stochastic error terms), a statistical relationship between an observed number of trips on a mode and the variables of that list would be neither the demand nor the supply relationship but a mixture of the two from which neither could be "unscrambled", to use a fitting expression (Ref. 2). The resulting "salad" relationship would not even be a reduced form equation.

It might be so difficult in practice to identify distinct supply and demand equations in the car trip market that one may be

condemned to use an aggregate "salad" relationship which one might call a "derived demand" equation and formulate as

$$(c_a \ v_a)^{dd} = T_a^{dd} = f^{dd}(K_{1a}, K_{2a}, P_a, K_{1b}, K_{2b}, P_b, OTHER_{da}, OTHER_{sa}), \quad (41)$$

and the parameters of which have no unambiguous interpretation. It is fortunate that demand and supply relationships for modes other than the car are probably overidentified because quantities supply and demanded typically depend on different lists of variables in the time periods over which the observation are defined.

Simultaneity and pure structural relationships. But, even with overidentified equations, consistent estimation of the parameters of an equation might be difficult. As an example, note that, under regulated pricing, the service level variables K_1 and K_2 are determined simultaneously by the behaviour of consumers and by that of producers. If, in a shuttle service, buses are added immediately when passenger queues are formed at the gate, a regression of the observed number of trips on service frequency and other determinants of demand would not yield a pure demand relationship because, with aggregate data, one would expect high frequencies to be correlated with high demands even if passengers were indifferent to waiting time. There are many situations in which line, area or time of day service levels are increased in anticipation of higher demands. Similarly, if higher ridership slows down vehicles, high travel times might be associated with high ridership in spite of passengers' dislike of in-vehicle time. Estimation procedures have to differentiate in demand equation parameter estimates between what reflects the suppliers' behaviour and what reflects the consumers'. Removing the simultaneous equations bias arising from supply-demand interactions may change the habitual expectations concerning money and time price demand elasticities estimated by limited-information techniques. Similar problems arise in estimating supply functions as soon as a measure of expected demand is used among the regressors. And even if the suppliers take no formal account of expected demand, the error term e_s of the supply equation is likely to be correlated with the error term e_d of the demand equation, a situation which reintroduces the simultaneous equations bias for different reasons.

Other stochastic specification problems. In most models, the size of the dependent variable varies considerably. This means that the error terms, particularly of the demand equations, are not likely to have a constant variance. In addition to being heteroskedastic, the error terms for different zones are likely to be correlated both spatially and over time. These forms of nonsphericity of the variance-covariance matrices of the error

terms of each equation are not invariant with respect to the functional form used to estimate the equation (Ref. 29) and should ideally be estimated jointly with the optimal form of the demand or supply function.

Within the appropriate time periods for which data are gathered, the existence of market clearing (in at least a weak sense), the identifiability of presumed relationship of interest, the explicitation of simultaneity in the structure and the correct specification of the stochastic error terms are essential requirements to the formulation and consistent estimation of travel supply and demand models. These conditions are important whether the presumed demand (or supply) relationship at hand results from a direct intuitive specification, from a formal link with utility theory or from analogy with a physical process (classical or relativistic, deterministic or stochastic) of any sort.

NETWORK EQUILIBRIUM

Reconsider the "derived demand" equation for car trips and the demand equation for bus trips of the previous section:

$$v_{ij_a}^{dd} = f_a^{dd}(K_{ij_a}, K_{ij_b}, OTHER_{ij_{da}}, OTHER_{ij_{sa}}) \quad (42)$$

$$c_b v_{ij_b}^d = f_b^d(K_{ij_a}, K_{ij_b}, OTHER_{ij_{db}}) \quad (43)$$

in which we have reintroduced i, j subscripts and written $K_{ij_a} = (P_{ij_a} \ K_{ij_{1a}} \ K_{ij_{2a}})$ and $K_{ij_b} = (\tilde{P}_{ij_b} \ K_{ij_{1b}} \ K_{ij_{2b}})$. In that section, K_{ij_a} and K_{ij_b} were assumed to be unique vectors of money and time prices. The assumption was apparently required to obtain, from appropriately linear parts of demand equations, a vector of parameters $\pi_m = (\pi_{Km} \ \pi_{other\ m})$ the elements of which are respectively associated with $K_{ij_m} = (K_{ij_a} \ K_{ij_b})$ and with $OTHER_m$ in the m^{th} equation. In this section, we shall reverse the situation and examine the determination of the K_{ij_m} vectors given the existence of demand parameter vectors π_m ; we shall then conclude with a few words on the problem* of the joint determination of π_m and K_{ij_m} .

The Assumption of Generalized Itinerary Cost Minimization

The determination of the vectors K_{ij_m} is not straightforward

*This difficulty was pointed out by Michael Florian.

because individual car drivers and transit passengers are in general faced with many possible itineraries between any two points; in fact, we will note that such vectors need not exist as long as a unique weighted sum of their components exists: that sum can be compatible with a great number of different itineraries between the same two points.

Indeed a natural approach to the explanation of how itineraries are chosen is to consider travellers as producers of their itineraries and to make hypotheses concerning their behaviour. And one such hypothesis is that the π_{km} parameters are "prices" used to weight money and time inputs required to go from i to j and that, as any cost minimizing entrepreneur of the standard theory of the firm, the traveller chooses the itinerary r which minimizes his total generalized cost incurred on network links l , or

$$GIC_{ij_a} = (\pi_{ka} K_{ij_a}^r) \text{ and } GIC_{ij_b} = (\pi_{kb} K_{ij_b}^r), \text{ where ' denotes}$$

transpose, $K_{ij}^r = \sum_{l(ij)r} \delta_{l(ij)r} K_l$ and $\delta_{l(ij)r}$ equals one if the link is chosen and zero otherwise. The way in which the sets of feasible itineraries or input combinations $\{K_{ij_a}^r\}$ and $\{K_{ij_b}^r\}$ arise is very complex.

Consider car trips. The set of feasible input combinations $\{K_{ij_a}^r\}$ depends on the set of link input requirements $\{K_{l_a}\}$ which are functions of physical (length and number of lanes, design, etc.), regulatory (tolls, signals, police, etc.) and other link factors N_l and of vehicle usage on the link V_{l_a} :

$$\{K_{ij_a}^r\} = F_{ij_a} \{K_{l_a}\}, \text{ where } K_{l_a} = f_{l_a}(N_l, V_{l_a}). \quad (44)$$

For given values of N_l and V_{l_a} , $\{K_{ij_a}^r\}$ is the set of all possible itineraries from i to j .

If cars are the sole users of a fixed supply of roads, Equation (44) is an appropriate description of input combination possibilities. A public transit authority, however, determines its link characteristics not only in relation to N_l and V_{l_b} but also, as

we saw in the previous section, in relation to its objectives, subsidies, etc., on each line and we may write

$$\{K_{ij_b}^r\} = F_{ij_b} \{K_{l_b}\}, \text{ where } K_{l_b} = f_{l_b} [f_{l_b}^s, f_{l_b}, (N_l, V_{l_b})] \quad (45)$$

is simply a link (or line) version of Equation (38) with endogenous link speeds; in (45) itineraries also consist of sets of links. One could envisage grouping Equations (44) and (45) to provide a full list of itineraries by mode when buses and cars interact on the road and even explaining the government's decisions on N_k in a general explanation of the supply of feasible input combinations*.

Given input "prices" π_{km} and trip demands by mode among all pairs of points $[v_{ij_a}^{dd}]$ and $[c_b T_{ij_b}^d]$, assignment algorithms derive specific values for GIC_{ij_a} and GIC_{ij_b} from the feasible sets $\{K_{ij_a}^r\}$ and $\{K_{ij_b}^r\}$: for $m = a$ and/or b , they effectuate

$$\begin{bmatrix} GIC_{ij_a} \\ GIC_{ij_b} \end{bmatrix}_m + \left(\pi_{km}, \{K_{ij_a}^r\}, \{K_{ij_b}^r\}, [v_{ij_a}^{dd}], [c_b v_{ij_b}^d] \right). \quad (46)$$

Obtaining the set of itinerary costs which minimize user itinerary costs does not mean that a unique vector is chosen for each specific pair of points: it only means that the generalized itinerary cost of all chosen itineraries between these points is unique. This is possible because it is not only the inputs which matter (distance, time, money tolls, or other physical performance or economic descriptors, such as risk of accident) but also the way in which they are perceived, weighted or transformed into a generalized cost which imports. Although road assignment procedures (Refs. 22 and 23) have a general validity, they are mostly used only with the travel time component of K_{ij_a} ; neither

are transit assignment procedures often used with more than one element of K_{ij_b} . Elsewhere (Ref. 15) we have said about the au-

tomobile part of (46): "the computation of the cheapest input requirements for each output level is what we may call the production equilibrium problem". We now believe that this terminology was slightly incorrect because standard production problems** minimize the total cost of producing a certain output for a system as a whole whereas, in this case, individual (user) itinerary costs are minimized: this means that, in the presence of congestion, where system and user solutions differ, the user-optimal

*The idea of a general presentation of vectors of itinerary input possibilities embodying Equations (44) and (45) as special cases is Michael Florian's who also suggested calling these vectors performance vectors.

**As was pointed out by Michael Florian.

solution is more akin to a decentralized production problem in which each unit (the origin-destination pair) is subject to decreasing returns to scale ($\partial GIC_{ij_m} / \partial V_{ij_m} > 0$) and produces externalities ($\partial GIC_{i'j'_m} / \partial V_{ij_m} \neq 0$) because each unit's costs depend on other units' output levels. In Hicksian terminology*, the street network is a set of techniques: it consists of a group of linked small plants or machines used simultaneously to produce different outputs. The transit network is also a set of techniques among which travellers choose with the help of their input "prices"; modifications of the network change the set of available techniques or, loosely speaking, the production "function".

The Network Equilibrium Problem

In reality it is not sufficient to consider trip demands as given and to find the GIC's for those demands: changes in GIC's feed-back on trip demands. If we write demand functions (42) and (43) for instance as

$$V_{ij_a}^{dd} = GIC_{ij_a} + GIC_{ij_b} + \pi_{other} OTHER_{ij_a} + e_{ij_a}^{dd}, \quad (47)$$

$$c_d V_{ij_b}^d = GIC_{ij_a} + GIC_{ij_b} + \pi_{other} OTHER_{ij_b} + e_{ij_b}^d,$$

the network equilibrium problem is to find values ($GIC_{ij_a}^*$, $GIC_{ij_b}^*$) and $V_{ij_a}^{dd*}$ and $c_b V_{ij_b}^d$ which solve (46) and (47) simultaneously when all other variables and parameters are held constant. (Refs. 11 and 12) have recently shown that, given demand parameters $\pi_m = (\pi_{km} \pi_{other m})$, it is possible to compute, for certain classes of demand functions, such network equilibria even if buses and cars interact on the common infrastructure.

If stochastic assignment procedures were ever used, one could envisage substituting the mean or expected value of GIC's for deterministic values in the demand functions.

Implications

Linearity in the form of the generalized cost function. Instead of using a weighted sum of variables which enter linearly in the generalized cost function, $GIC_{ij} = \sum_k \gamma_k K_{ij_k}$, one might wish to use a measure of generalized cost derived from a more general

*As Ake Andersson reminded us.

form of the demand function, $GIC_{ij} = \sum_k \gamma_k K_{ijk}^{(\lambda)}$, where λ denotes

the standard Box-Cox (Ref. 4) transformation function. The growing use of these transformations in model estimation will imply new problems in the computation of user optimal paths in congested networks: further research is clearly needed to work out reasonable network equilibrium procedures in the context of generalized functional forms of demand equations (e.g. Ref. 18).

Market equilibrium and network equilibrium. The distinction between supply-demand equilibrium and network equilibrium can be expressed as follows: empty buses (excess supply) and having to wait for more than one bus to obtain space (frequent excess demand) are perfectly compatible with the existence of a simultaneous solution for (46) and (47). It is difficult to visualize in the case of excess demand because the solution is obtained through queues which cannot be measured by a simple transformation of data from vehicle operations. When variations in

$v_{ij_a}^{dd}$ and in $c_b v_{ij_b}^d$ cause changes in link characteristics

(because cars cause congestion or bus passengers slow down service) which imply changes in generalized itinerary costs, a new network equilibrium has to be computed. As such, that equilibrium is in general perfectly consistent with many degrees of aggregate excess supply in the car market (low occupancy rates) and in the bus market (empty buses). Shifts of network supply or demand variables (legal car driving age, transit management objectives, etc.), will imply a new network equilibrium and an entirely different market disequilibrium. This distinction between market and network equilibrium could be made more spatially precise (link specific excess demand on buses in consistent with system-wide excess supply, and so on) and interesting situations could be derived from the fact that parameter vectors " K_m " in

general differ among mode equations, but we shall instead belabour another implication of network equilibrium procedures, namely the natural separability of demand estimation and network equilibrium procedures. We shall drop i, j subscripts, writing $K_m = (K_a K_b)$.

Demand estimation and the true itinerary input vector. In the section on market equilibrium, we examined the problem of estimation of $\pi_m = (\pi_{K_m} \pi_{other\ m})$, given K_m . In this section we have looked at the determination of K_m , given π_m . Can they be jointly determined? Faced with this difficulty, as he was estimating the demand function from observed link flows, Wills (Ref. 28) found that, using uncongested assignment procedures of estimated origin-destination demands, he could determine K_m and π_{K_m} separately from $\pi_{other\ m}$ by iterating successively on π_m and K_m , for constant demands by origin-destination pair, and on π_m , for constant input vectors K_m . This result may have been due to the

fact that, with O-D demands of strongly varying sizes (including a few very large ones), K_m was in fact nearly constant or insensitive to small changes in π_{Km} in a certain domain. He did not, however, reach convergence when he tried to determine K_m and π_m jointly. This result may have been attributable to the fact that this assignment procedure in the road network caused a diversion on many paths, thus implying a different value for some K_a 's at each iteration; it may also have been caused by changes in all-or-nothing itineraries on buses when fare, departure frequency and travel time elements of K_b were considered in the assignment. Had only one element been used in all-or-nothing assignment procedures, there would have been no clear reason to cycle. It is interesting to conjecture on what would have happened if he had allowed congestion to occur: convergence might not have been reached, even for constant pair demands. The absence of a unique solution might have been even more apparent than with uncongested networks. Further research in this area is needed: note for instance that, if the criterion function which is minimized in the estimation of π_m is formulated in terms of errors on pairs rather than on links, the problem appears to be more difficult. Such increased difficulty might arise because, for given values of $\pi_{\text{other } m}$, a greater number of combinations of π_{Km} and K_m would be compatible with a certain value of the criterion function; in consequence, the indeterminacy noted by Wills would be more likely to appear, even under congested assignment procedures.

From our point of view, this problem of joint determination of π_m and K_m does not radically change the problem of estimation of the demand function because consistent estimation of π_m does not require actual observations on the true value of K_m . All that is needed for consistency is an initial vector of instruments $\hat{K}_m$ correlated with the true and unknown K_m and asymptotically independent of the error terms in (47). As a first approximation it appears that nothing could be gained by subsequently iterating on $\hat{K}_m$ and π_m but further research might show that a more efficient $\hat{K}_m$ might thus be obtained. Such a procedure amounts to taking into account an error of observation on K_m and does not basically differ from those we referred to in our earlier discussion of simultaneous equations biases and errors of observation on service level variables. The procedure implies a separation of demand estimation from network equilibrium procedures in which the consistent estimates of $\pi_m = (\pi_{Km} \ \pi_{\text{other } m})$ obtained from the demand estimation are subsequently used in the network equilibrium problem.

DYNAMIC EQUILIBRIUM

Changes in infrastructure in one period lead to changes in generalized itinerary costs over time; similarly, changes in other determinants of demand or of the behaviour of suppliers take time to make themselves felt. Demand function (1) can in fact be written as

$$T_t^d = f^d \left(\sum_n PD_{T_{t-n}}, \sum_n S_{T_{t-n}}, P_{v_t}, P_{f_t}, A_{f_t} \right) + e_t^d,$$

$$n = 0, 1, \dots, (48)$$

and it is part and parcel of the explanation of observed travel behaviour to determine these time paths which depend on the costs of acquiring information and other market imperfections. In fact all of the equilibria we have studied in the previous section may or may not occur within the time period over which observations are defined.

Although little is known in general about dynamic equilibrium, it should not be thought that these time paths are obvious. For instance, recent calculations (Ref. 17) comparing fare and waiting time transit demand elasticities in Montreal brought out a variable-specific response and a market asymmetry: as shown in Table 2, increases in adult fares resulted in lower passenger losses over two periods than over one period (e.g. some riders came back) but increases in waiting time resulted in higher losses over two periods than over one period (e.g. losses are permanent); the opposite appeared to be true for schoolchildren but the results for that market were statistically less sturdy than the results for the adult market.

TABLE 2 One-Month and Two-Month Transit Demand Elasticities and t-Statistics for Montreal Adults and Schoolchildren*

	Adults		Schoolchildren	
	1-period	2-period	1-period	2-period
Net fare elasticity	.16 (-4.94)	.14	.44 (-3.64)	.45
Net waiting time elasticity	.36 (-3.77)	.48	.21 (-0.75)	.05

Further analysis of the establishment of travel equilibrium over time could reveal the existence of analogous differences across income groups, trip purposes, and other market segmentation criteria. Models should clearly incorporate the dynamic equilibrium

*Further details supplied in Appendix B. Results for adults differ from those of Table 1 due to updating of sample base.

features which transportation system managers have often felt relevant to policy decisions. We have noted (Ref. 14) that the transit suppliers can react to past values of demand and adjust service levels with a lag which implies that market disequilibria can take time to disappear and should make a difference to model formulation and estimation.

CONCLUSION

We have chosen paradigm equations to center our attention on six notions of equilibrium and point out some of their implications for the explanation of travel behaviour. We could have chosen other paradigms to discuss the same or other issues. We have not developed important topics such as that of the relationship between car seat trips and car vehicle trips or as the problem of network aggregation which introduces new errors of observation on the money and time itinerary inputs; we have also abstained from discussing the specification error arising from the standard consumer homogeneity assumption of demand analysis and from the use of variables to account for the actual heterogeneity of samples. These topics and others, such as levels of aggregation of models, are important to travel modelling but are perhaps less closely connected to notions of equilibrium than those we have selected in the hope of eliciting further discussions of these issues and of their methodological implications.

APPENDIX A

The joint supply formulation of the dependent variable arises out of the simple idea that if trip characteristics constitute the travel good, then purchases of these characteristics can be used as the dependent variable of travel demand functions. The models we compared were based on the following structure:

$$y_t = x_t \beta + u_t, \quad (49)$$

$$u_t = \rho_1 u_{t-1} + \rho_{12} u_{t-12} + e_t, \quad (50)$$

where all observations are monthly observations and e_t is a random vector of independent and identically distributed error terms with zero constant variance σ^2 . The matrix of observation contains service levels of public transit and competing modes, comfort levels of public transit and competing modes, comfort levels, activity levels and other variables. Under the standard format, the vector y_t is simply the number of adult passengers and, under the joint supply format, it is $\alpha_1 K_{1t} y_t + \alpha_2 K_{2t} y_t + \alpha_3 K_{3t} y_t$, where $\alpha_1 + \alpha_2 + \alpha_3 = 1$. In the standard case the system consisting of Equations (49) and (50) is estimated by a procedure (Ref. 14) which yields efficient estimates of the parameters β , ρ_1 , ρ_{12} and σ^2 . In the joint consumption format all but one of the

elements of the dependent variable are put on the right-hand side of Equation (49) and the presence of these endogeneous variables as regressors biases the parameters estimates obtained from the same limited-information technique. The results were invariant with respect to the normalization procedure used. Table 1 presented some elasticities. Further details can be found in (Ref. 3).

APPENDIX B

The following equations for the schoolchildren market can be added to the adult market equations (49) and (50)

$$y_t = X_t \beta + u_t, \quad (51)$$

$$u_t = \rho_1 u_{t-1} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_{12} u_{t-12} + e_t \quad (52)$$

and both systems can be estimated together using a generalized and iterated version of Park's method (Ref. 24). To test for lags in speed of adjustment of adults and schoolchildren to fare and waiting time changes, lagged values of these variables were added to the regressor lists of Equations (49) and (51). In Table 2 t-statistics are given for the fare and waiting time elasticities in the absence of adjustment lags. No such statistics are given for the 2-period results because these statistics exist only for each of the variables the parameters of which were combined to yield the 2-period adjustment results. Further details and results are available in (Ref. 17) which also contains a summary of market forecasts performed with estimates obtained under various estimation techniques.

BIBLIOGRAPHY

- (1) Barten, A.P., Evidence on the Slutsky conditions for demand equations, Review of Economics and Statistics, 49, p. 77-84 (1967).
- (2) Baumol, W.C., (1965) Economic Theory and Operations Analysis, Prentice-Hall, New Jersey, ch. 10.
- (3) Bilodeau, D., La formulation des équations de demande de transport: un problème de consommation conjointe, publication #74, Centre de recherche sur les transports, Université de Montréal (1977).
- (4) Box, G.E.P. and Cox, D.R., An Analysis of Transformations, Journal of the Royal Statistical Society B 211, p. 211-243 (1964).
- (5) Charles River Associates, A Disaggregated Behavioral Model of Urban Travel Demand, Report No. CRA-156-2, Cambridge, Mass. (1972).

- (6) De Serpa, A.C., A Theory of the Economics of Time, The Economic Journal, 81 p, 828-846 (1971),
- (7) Domencich, T.A., Kraft, G. and Valette, J.P., Estimation of Urban Passenger Travel Behavior: An Economic Demand Model, Highway Research Record 238 (1968).
- (8) Domencich, T.A. and McFadden, D. (1975). Urban Travel Demand, North Holland, Amsterdam.
- (9) Fair, R.C. and Jaffee, D.M., Methods of Estimation for Markets in Disequilibrium, Econometrica, 40, p. 497-514 (1972).
- (10) Fairhurst, M.H. and Morris, P.J., Variations in the Demand for Bus and Rail Travel Up to 1974, Economic Research Report R 210, London Transport, London (1975).
- (11) Florian, M., A Traffic Equilibrium Model of Travel by Car and Public Transit Modes, publication # 32, Centre de recherche sur les transports, Université de Montréal (1975).
- (12) Florian, M. Urban Travel Demand Models and Multi-Modal Traffic Equilibrium, Proceedings, Seventeenth Annual Meeting, Transportation Research Forum, p. 184-190 (1976).
- (13) Gaudry, M., The Demand for Public Transit in Montreal and its Implications for Transportation Planning and Cost-Benefit Analysis, Ph. D. dissertation, Princeton University (1973).
- (14) Gaudry, M., An Aggregate Time-Series Analysis of Urban Transit Demand: The Montreal Case, Transportation Research 9, p. 249-258 (1975).
- (15) Gaudry, M., A Note on the Economic Interpretation of Delay Functions in Assignment Problems, Lecture Notes in Economics and Mathematical Systems 118, Springer-Verlag, New York, p. 368-382 (1976a).
- (16) Gaudry, M., Passenger Demand Choice of Mode of Travel: Some Dimensions of Models, Report E.P. 210 1-G-50, Ministry of State for Urban Affairs, Ottawa (1976b).
- (17) Gaudry, M., Seemingly Unrelated Urban Travel Demands, publication # 61, Centre de recherche sur les transports, Université de Montréal (1977).
- (18) Gaudry, M. and Wills, M., Estimating the Functional Form of Travel Demand Models, publication # 63, Centre de recherche sur les transports, Université de Montréal (1977).
- (19) Goldberger, A.S., Functional Form and Utility: A Review of Consumer Demand Theory, Social Systems Research Institute Workshop Paper 6703, University of Wisconsin (1967).

- (20) Huang, T.J., An Econometric Analysis of the Determinants of Demand for Local Transit in Edmonton, 1961-1970, Report No. 1, Finance Department, City of Edmonton (1973).
- (21) Kresge, D.T. and Roberts, P.O. et al. (1971) Systems Analysis and Simulation Models, in Meyer, J.R., ed.: Techniques of Transport Planning II, The Brookings Institution, Washington, D.C.
- (22) Nguyen, S., An Algorithm for the Traffic Assignment Problem, Transportation Science 8, No. 3, p. 203-216 (1974a).
- (23) Nguyen, S., Une approche unifiée des méthodes d'équilibre pour l'affectation du trafic, Ph.D. dissertation, publication #171, département d'informatique, Université de Montréal (1974b).
- (24) Parks, R.W., Efficient Estimation of a system of Regression Equations when disturbances are both Serially and Contemporaneously Correlated, J.A.S.A. 62, p. 500-509 (1967).
- (25) Quandt, R.E., Testing Hypotheses in Disequilibrium Models, Econometric Research Program Memo # 197, Princeton University (1976a).
- (26) Quandt, R.E., Maximum Likelihood Estimation of Disequilibrium Models, Econometric Research Program Memo #198, Princeton University (1976b).
- (27) Samuelson, P.A., Contrast Between Welfare Conditions for Joint Supply and for Public Goods, Review of Economics and Statistics II, p. 26-30 (1969).
- (28) Wills, M.J., Linear and Non-Linear Estimators of the O.D. Matrix, Ph. D. dissertation, Department of Geography, University of British Columbia, Vancouver (forthcoming) chapter 6.
- (29) Zarembka, P., Transformation of Variables in Econometrics, Ch. 3 in Frontiers in Econometrics, edited by P. Zarembka, Academic Press (1974).
- (30) Zellner, A., An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, J.A.S.A. 57, p. 348-368 (1962).